Journal of Multidisciplinary Healthcare

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Development and Validation of a Clinical Prediction Model for Growth Hormone Deficiency in Children with Short Stature: A Retrospective Study in China

Mali Li, Chao Liu, Yuan Yang, Shuwen Hu, Jia Li, Shichao Qiu, Zhihua Wang

Department of Endocrinology, Genetics and Metabolism, Xi'an Children's Hospital, Xi'an, Shaanxi, People's Republic of China

Correspondence: Zhihua Wang, Department of Endocrinology, Genetics and Metabolism, Xi'an Children's Hospital, Xi'an, Shaanxi, People's Republic of China, Email xasetyy@126.com

Background: A multitude of congenital and acquired conditions can result in short stature, each with distinctive clinical presentations and treatment options. We aimed to develop and validate a prediction model to identify GHD among children with short stature using clinical and laboratory parameters.

Methods: This retrospective observational study included 1120 children with short stature from a hospital in China. The data were randomly split into a derivation set and a validation set. Features were selected based on clinical relevance and statistical significance to construct a multivariate logistic regression model in the derivation set. Discrimination, calibration, and prediction accuracy were evaluated on both sets.

Results: Of the 1120 children, 278 (25%) were diagnosed with GHD, 694 (62%) were male, and the mean age was 6.97 ± 2.97 years. The derivation set comprises 785 (70%) children. The model incorporates four predictors: age (OR=0.761; 95% CI 0.660, 0.873), delayed bone age (OR=1.841; 95% CI 1.365, 2.537), IGF-1 SDS (OR=0.148; 95% CI 0.095, 0.220), and IGF-1/IGFBP-3 ratio (OR=0.901; 95% CI 0.870, 0.930). The model exhibits good discriminative ability, with an AUC of 0.952 (0.937, 0.967) in the derivation set and 0.950 (0.927, 0.973) in the validation set. Furthermore, it shows high accuracy with sensitivity and specificity of 0.895 in the derivation set, which was 0.946 and 0.851 in the validation set. The model also demonstrates reliable calibration.

Conclusion: We have developed a prediction model for accurate screening of GHD in children with short stature.

Keywords: growth hormone deficiency, short stature, prediction model, insulin-like growth factor 1, insulin-like growth factor binding protein 3, bone age

Introduction

Short stature is characterized by an adult height of more than two standard deviations (SDs) below the average height for a given age and sex within a specific population, corresponding to the shortest 2.3% of individuals in that population. It is associated with impaired intellectual and physical dysfunction[1] and poor quality of life.[2]

The growth hormone (GH)/insulin-like growth factor-I (IGF-I) axis was believed to have the primary role in short stature. However, recent research indicates that this is only one of multiple regulatory systems that oversee chondrogenesis in the growth plate, which is the biological mechanism responsible for height gain. Normal growth in children relies not only on GH and IGF-I but also on various hormones, paracrine factors, extracellular matrix molecules, and intracellular proteins that control the growth plate chondrocytes.[3] Short stature can therefore be caused by a multitude of congenital and acquired conditions, such as growth hormone deficiency (GHD), idiopathic short stature (ISS), children born small for gestational age (SGA), and various genetic syndromes.[4]

Received: 16 April 2025
Accepted: 1 August 2025
Published: 5 September 2025

GHD is a condition characterized by inadequate secretion of GH from the pituitary gland and is the most prevalent abnormality within the GH/IGF-1 axis. SGA is defined as birth weight and/or length at least 2 SDs below the average for gestational age. Within the SGA population, approximately 10 to 15% of children do not achieve catch-up growth and exhibit a reduced adult height.[5] Many genetic syndromes, such as Noonan syndrome, Turner syndrome, and Prader-Willi syndrome, are also linked with hormonal imbalances, affecting an individual's growth and contributing to short stature.[6] The diagnosis of ISS is made after other conditions causing short stature have been ruled out. ISS encompasses a diverse group of children with varying phenotypes and genotypes and can be divided into familial and non-familial categories. Delayed skeletal maturation and onset of puberty may also be observed.[7,8]

It is essential to make an accurate diagnosis of GHD in children with short stature. Unlike short stature resulting from other etiologies, GHD can be associated with concurrent deficiencies in pituitary hormones or central nervous system tumors.[9] GHD has also been reported to respond more positively to GH therapy, though clinical management and outcomes differ for patients with active oncological disease.[10] The gold standard for diagnosing GHD is the GH stimulation test,[10] as it assesses the response of the hypothalamus and pituitary gland to different stimuli. However, the diagnosis may be challenging in clinical settings. Firstly, due to the frequent blood sampling involved, children and their families may have a low tolerance for this test. Moreover, the test can lead to adverse reactions and risks.[11] Researchers have recommended that GH stimulation tests should only be reserved for individuals where the results are the ultimate and decisive factor for intervention.[12]

Markers have been proposed for diagnosing GHD. Clinical features, including serum IGF-1, insulin-like growth factor binding protein-3 (IGFBP-3), and the IGF-1 to IGFBP-3 ratio, have been previously recognized as useful diagnostic markers for diagnosing GHD and can be utilized as auxiliary diagnostic markers for GH stimulation tests.[13–17] According to a meta-analysis of 12 studies, the summary receiver operating characteristic curve (ROC) of IGF-1 was 0.78, and the summary ROC of IGFBP-3 was 0.80.[18] Magnetic resonance imaging (MRI) features, including pituitary volumes, have also been proven useful.[19,20]

Despite the availability of the GH stimulation test and various auxiliary markers, current diagnostic approaches for GHD have significant limitations. The GH stimulation test, while considered the gold standard, is invasive, time-consuming, carries risks, and has low tolerance among patients and families, making it impractical for widespread use, especially in primary care or low-resource settings. While markers like IGF-1 and IGFBP-3 are helpful, their individual predictive performance for GHD diagnosis is only moderate. Furthermore, most existing research has focused on univariate analyses of these markers, which are insufficient because they fail to account for the complex interplay and confounding factors among multiple clinical and biochemical variables relevant to GHD diagnosis.

Only a few studies have reported prediction models for diagnosing GHD, and their performances are poor to moderate, often relying on advanced imaging or non-routine parameters that restrict their applicability in primary care or low-resource settings. A study developed a prediction model for GHD in a cohort including 1496 Chinese short stature children.[21] The predictors identified included age, body mass index (BMI), ALT, IGF-1, and IGFBP-3. The model has poor discrimination, with an area under the curve (AUC) of 0.593. A study including 362 short stature children in China developed a prediction model for GHD using MRI texture features, which has an AUC of 0.852.[22] What is lacking in the current diagnostic approaches is a non-invasive, accessible, and accurate prediction model that can effectively screen for GHD before resorting to the burdensome GH stimulation test.

This study directly addresses these gaps by developing and validating a new prediction model for GHD. Our model utilizes readily available clinical features to accurately distinguish between GHD and other causes of short stature in children, thereby reducing the need for preliminary GH stimulation tests. This approach aims to provide a more practical and widely applicable diagnostic tool for GHD, particularly in settings where advanced imaging or complex tests are not feasible.

Materials and Methods

Study Population

Patients under the age of 18 admitted for short stature at the pediatric endocrinology department of a tertiary hospital were retrospectively selected from hospital records. Short stature was defined as having a height below −2 SD, or the third

percentile of the normal growth curve for children of the same age and gender.[23] GHD was further diagnosed using the prescribed criteria, as follows: (1) height velocity of no more than 5 cm per year; (2) serum GH peak of less than 10 μg/L in two separate GH stimulation tests using arginine or clonidine as stimulants;[24] (3) delayed bone age compared to chronological age; (4) serum IGF-1 level below normal; (5) symmetric short stature and infantile facial features; and (6) normal intellectual development. Diagnosis criteria for other causes of short stature are elaborated in supplementary methods.

Patients were included if their cause of short stature is among the most frequent causes of GHD, ISS, SGA, and genetic syndromes. Other patients were excluded from the study due to the low incidence rates. Children meeting any of the following criteria were further excluded from the study: (1) missing data of age, sex, height, IGF-1 or IGFBP-3 levels, or bone age; (2) BMI falling beyond the 15th to 85th percentile range for children of the same age and sex; (3) presence of distinct facial features, body shapes, skeletal deformities, or chondrodysplasia; (4) presence of chronic systemic diseases such as malnutrition, heart disease, liver disease, kidney failure, or lung disease; and (5) prior exposure to GH therapy.

The study was approved by the Ethics Committee of Xi'an Children's Hospital. Due to the retrospective nature of the study and the use of fully anonymized data, a waiver of informed consent was obtained. The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Data Preparation

To ensure robust data protection, all patient data were de-identified at the source before their transfer for analysis, removing any direct or indirect identifiers. The anonymized dataset was securely stored on password-protected institutional servers, with access restricted exclusively to authorized members of the study's research team. Strict data governance protocols were adhered to throughout the study, in compliance with data privacy regulations and ethical guidelines for pediatric research, ensuring the confidentiality and integrity of all patient information. Trained investigators reviewed patient charts using a standardized case report form. Deidentified data, including demographic characteristics and clinical features, were collected.

The height standard deviation score (SDS) was calculated as (height at evaluation time point – average height of children of same age and sex)/(SD of height of children of same age and sex) using Chinese reference values.[23]

The SDS of serum IGF-1 was calculated based on data and the method from a study in China[25] using the skewness (L), median (M), and coefficient of variation (S) parameters using the formula below, where Y_{IGF-1} denotes the measured serum IGF-1 level.

$$\text{IGF - 1 SDS} = \frac{(Y_{IGF-1}/M)^{L} - 1}{L \times S}$$

The delayed bone age was calculated as the difference between the chronological age and the bone age.

Growth Hormone Stimulation Test

The stimulation test was initiated at 6:00 after an overnight fasting. Blood samples were collected at 0, 30, 60, 90, and 120 minutes following stimulus administration. Clonidine (4 μg/kg, capped at 150 μg) and 10% arginine (0.5 g/kg, capped at 30 g) were administered as stimulating agents. Serum GH levels were determined using chemiluminescence immunoassay on the IMMULITE® 2000 Immunoassay System (Siemens, Germany). Assays were performed according to the manufacturer's instructions.

Statistical Analysis

Missing data was rare, as shown in Table S1. Missing data were introduced during the calculation of IGF-1 SDS due to zero values of L, and the samples were removed from further analysis. The data was randomly split into two analysis sets: a derivation set including 785 (70%) patients and a validation set including 335 (30%) patients.

Candidate features for the multivariate prediction model were identified through a two-step process integrating clinical relevance and statistical association within the derivation set. First, among characteristics that are significantly different between GHD and non-GHD patients, variables deemed clinically relevant to short stature and GHD

diagnosis by clinical experts were identified. Subsequently, these clinically relevant variables were screened for statistical significance through univariate logistic regression, with the diagnosis of GHD as the dependent variable. Variables demonstrating a statistically significant association ($P < 0.05$) were then considered as candidate features. A prediction model was then constructed using multivariate logistic regression with the candidate features. Multicollinearity among the predictors was assessed using variance inflation factors (VIFs). Features were further removed if they were not statistically significant in the multivariate model, or if they demonstrated significant multicollinearity (VIF ≥ 5).

Model performance was evaluated in both analysis sets. ROC analysis was conducted to evaluate the discrimination of the model. Calibration was assessed using a visual calibration plot and the Hosmer-Lemeshow test, with the χ^2 value and *P*-value reported. Additionally, the Brier score was calculated to provide a more robust evaluation of model performance. The optimal threshold was selected by leveraging sensitivity and specificity. Under the selected threshold, model performance measures including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated.

All statistical analyses were performed using R (version 4.1.2). Continuous variables were presented as mean ± SD and compared using unpaired t-tests. Categorical variables were presented as count (%) and compared using χ^2 tests. A *P*-value of <0.05 was used as a statistically significant threshold.

Results

Characteristics of the Study Population

A total of 1120 children were included in the analysis, including 278 (25%) GHD patients, 650 (58%) ISS patients, 98 (9%) SGA patients, and 94 (8%) short stature children with genetic syndromes. Of all patients, 694 (62%) were male, and the mean age was 6.97 ± 2.97 years. Characteristics of the overall study population, GHD patients, and non-GHD patients are shown in Table 1. Comparative analysis revealed significant differences between GHD and non-GHD patients. Children with GHD were significantly older (7.76 ± 3.19 vs 6.71 ± 2.85 years; $P < 0.001$) and more likely to be male (72% vs 59%; $P < 0.001$). Auxological assessment demonstrated greater growth impairment in the GHD group, evidenced by lower height SDS (−2.65 ± 0.59 vs −2.52 ± 0.79; $P = 0.003$), greater bone age delay (2.58 ± 0.40 vs 1.95 ± 1.09 years; $P < 0.001$), and higher BMI (16.04 ± 1.83 vs 15.75 ± 2.58 kg/m²; $P = 0.041$). Biochemical profiling showed markedly lower IGF-1 SDS (−2.50 ± 0.94 vs −0.63 ± 1.25; $P < 0.001$) and IGF-1/IGFBP-3 ratio (21 ± 9 vs 39 ± 16; $P < 0.001$) in GHD patients, along with the expected attenuated GH peak response (5.4 ± 2.4 vs 13.3 ± 3.8 μg/L; $P < 0.001$). Notably, the GHD group had higher birth weights (3.18 ± 0.34 vs 2.98 ± 0.48 kg; $P < 0.001$) but no family history of short stature (0% vs 15.6%; $P < 0.001$), while maternal height was greater in GHD cases (157.2 ± 4.2 vs 156.5 ± 5.5 cm; $P = 0.026$). A detailed description of the patient characteristics stratified by the cause of short stature is shown in Table S1.

Model Development

The derivation set and the validation set comprise 785 (70%) and 335 (30%) patients, respectively. The proportion of GHD patients is 199 out of 785 (25%) in the derivation set and 79 out of 335 (24%) in the validation set.

Among characteristics that are significantly different between GHD and non-GHD patients, those deemed clinically useful are considered variables in the feature selection step. Univariate logistic regression was conducted using each variable as the independent variable and the diagnosis of GHD as the dependent variable. The odds ratio (OR) and the 95% confidence interval (CI) are shown in Figure 1. Among the variables, all are statistically significant except for BMI. IGF-1 and IGF-1 SDS have similar implications, and we included only IGF-1 SDS for further analysis. For the final multivariate model, all VIF values were below the threshold of 5 (age: 2.097; delayed bone age: 1.030; IGF-1 SDS: 1.828; and IGF-1/IGFBP-3 ratio: 1.189).

Model Specification

Features were further removed if they were not statistically significant in the multivariate model. The final multivariate logistic regression model parameters are shown in Table 2, which includes four predictors: age (OR=0.761; 95% CI

Table 1 Characteristics of the Study Population

Characteristic	Overall N = 1,120	GHD N = 278	Non-GHD N = 842	*P*-value
Etiology				<0.001
GHD	278 / 1,120 (25%)	278 / 278 (100%)	0 / 842 (0%)	
ISS	650 / 1,120 (58%)	0 / 278 (0%)	650 / 842 (77%)	
SGA	98 / 1,120 (8.8%)	0 / 278 (0%)	98 / 842 (12%)	
Genetic syndromes	94 / 1,120 (8.4%)	0 / 278 (0%)	94 / 842 (11%)	
Sex				<0.001
Female	426 / 1,120 (38%)	79 / 278 (28%)	347 / 842 (41%)	
Male	694 / 1,120 (62%)	199 / 278 (72%)	495 / 842 (59%)	
Age (years)	6.97 ± 2.97	7.76 ± 3.19	6.71 ± 2.85	<0.001
Age group				<0.001
< 3 years	31 / 1,120 (2.8%)	0 / 278 (0%)	31 / 842 (3.7%)	
3~5 years	478 / 1,120 (43%)	110 / 278 (40%)	368 / 842 (44%)	
6~12 years	534 / 1,120 (48%)	131 / 278 (47%)	403 / 842 (48%)	
≥ 12 years	77 / 1,120 (6.9%)	37 / 278 (13%)	40 / 842 (4.8%)	
Bone age (years)	4.88 ± 2.91	5.21 ± 3.13	4.77 ± 2.83	0.036
Delayed bone age (years)	2.10 ± 1.01	2.58 ± 0.40	1.95 ± 1.09	<0.001
Height (cm)	109 ± 16	113 ± 16	108 ± 16	<0.001
Height SDS	−2.55 ± 0.75	−2.65 ± 0.59	−2.52 ± 0.79	0.003
Weight (kg)	19 ± 7	21 ± 8	19 ± 6	<0.001
BMI (kg/m2)	15.82 ± 2.41	16.04 ± 1.83	15.75 ± 2.58	0.041
IGF-1 (ng/mL)	120 ± 64	69 ± 29	136 ± 64	<0.001
IGF-1 SDS	−1.07 ± 1.43	−2.50 ± 0.94	−0.63 ± 1.25	<0.001
IGFBP-3 (μg/mL)	3.87 ± 8.89	4.51 ± 17.71	3.66 ± 1.25	0.400
IGF-1/IGFBP-3 ratio (×1000)	34 ± 17	21 ± 9	39 ± 16	<0.001
Growth hormone peak (μg/L)	11.3 ± 4.9	5.4 ± 2.4	13.3 ± 3.8	<0.001
Premature birth	46 / 1,120 (4.1%)	7 / 278 (2.5%)	39 / 842 (4.6%)	0.12
Type of delivery				0.2
Cesarean section	411 / 1,120 (37%)	111 / 278 (40%)	300 / 842 (36%)	
Vaginal birth	709 / 1,120 (63%)	167 / 278 (60%)	542 / 842 (64%)	
Birth weight (kg)	3.03 ± 0.46	3.18 ± 0.34	2.98 ± 0.48	<0.001
Father's height (cm)	168.9 ± 8.2	168.1 ± 13.6	169.1 ± 5.4	0.200
Mother's height (cm)	156.6 ± 5.2	157.2 ± 4.2	156.5 ± 5.5	0.026
Family history of short stature	131 / 1,120 (12%)	0 / 278 (0%)	131 / 842 (16%)	<0.001
ACTH (pmol/L)	20 ± 17	20 ± 12	21 ± 18	0.600
Cortisol (μg/dL)	15.3 ± 5.3	15.5 ± 7.3	15.3 ± 4.4	0.700
Fasting insulin (mIU/L)	8.4 ± 5.1	8.5 ± 7.1	8.3 ± 4.3	0.800

Abbreviations: GHD, growth hormone deficiency; ISS, idiopathic short stature; SGA, small for gestational age; SDS, standard deviation score; BMI, body mass index; IGF-1, insulin-like growth factor-1; IGFBP-3, insulin-like growth factor binding protein 3; ACTH, adrenocorticotropic hormone.

0.660, 0.873), delayed bone age (OR=1.841; 95% CI 1.365, 2.537), IGF-1 SDS (OR=0.148; 95% CI 0.095, 0.220), and IGF-1/IGFBP-3 ratio (OR=0.901; 95% CI 0.870, 0.930). The optimal threshold was selected to be 0.282 by leveraging sensitivity and specificity (Figure S1).

Model Performance

Model performance was evaluated in the derivation set and the validation set, respectively. The model exhibits good discriminative ability, with an AUC of 0.952 (0.937, 0.967) in the derivation set and 0.950 (0.927, 0.973) in the validation set (Table 3 and Figure 2a). The calibration plot shows that the goodness-of-fit character for our model is satisfactory (Figure 2b), with Hosmer-Lemeshow *P*-values of 0.614 (χ^2 = 1.804) for the derivation set and 0.189 (χ^2 =

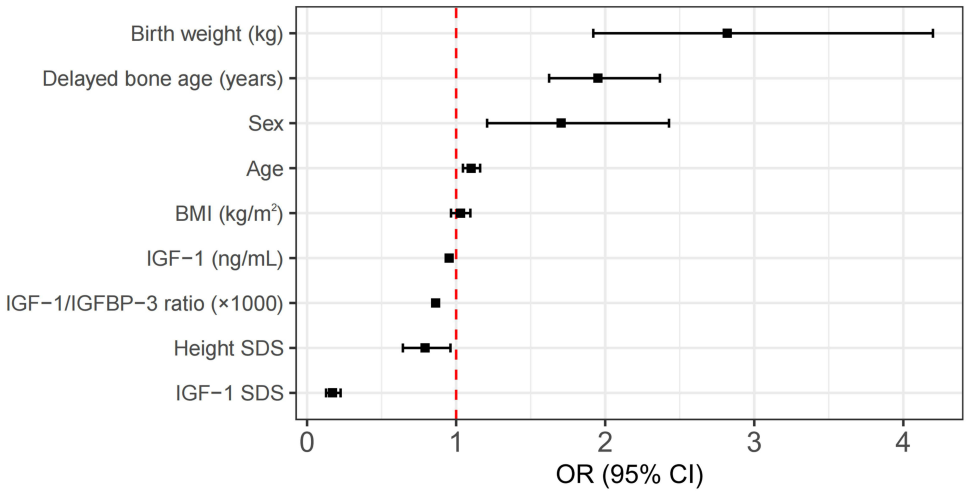

Figure 1 Univariate logistic regression result.
Abbreviations: SDS, standard deviation score; BMI, body mass index; IGF-1, insulin-like growth factor-1; IGFBP-3, insulin-like growth factor binding protein 3; OR, odds ratio; CI, confidence interval.

4.779) for the validation set. Additionally, the Brier score was 0.072 for the derivation set and 0.076 for the validation set. At the selected threshold, the model showed good performance in all prediction accuracy metrics in both analysis sets (Table 3). Specifically, the model demonstrates a sensitivity of 0.895 and a specificity of 0.895 in the derivation set, which was 0.946 and 0.851 in the validation set.

Discussion

In this study, we developed an accurate prediction model to screen for GHD in short stature children. We propose that this model be used as a primary screening tool for GHD in children presenting with short stature. Children who obtain positive GHD screening results via the model can undergo further verification through gold-standard GH stimulation tests.

Table 2 Multivariate Logistic Regression Model Parameters

Variable	OR (95% CI)	*P*-value
(Intercept)	0.393 (0.122, 1.234)	0.113
Age	0.761 (0.660, 0.873)	<0.001
Delayed bone age (years)	1.841 (1.365, 2.537)	<0.001
IGF-1 SDS	0.148 (0.095, 0.220)	<0.001
IGF-1/IGFBP-3 ratio (×1000)	0.901 (0.870, 0.930)	<0.001

Abbreviations: SDS, standard deviation score; IGF-1, insulin-like growth factor-1; IGFBP-3, insulin-like growth factor binding protein 3; OR, odds ratio; CI, confidence interval.

Table 3 Model Performance Metrics

Analysis Set	AUC (95% CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Derivation set	0.952 (0.937, 0.967)	0.895	0.895	0.895	0.730	0.964
Validation set	0.950 (0.927, 0.973)	0.873	0.946	0.851	0.654	0.981

Abbreviations: AUC, area under the curve; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value.

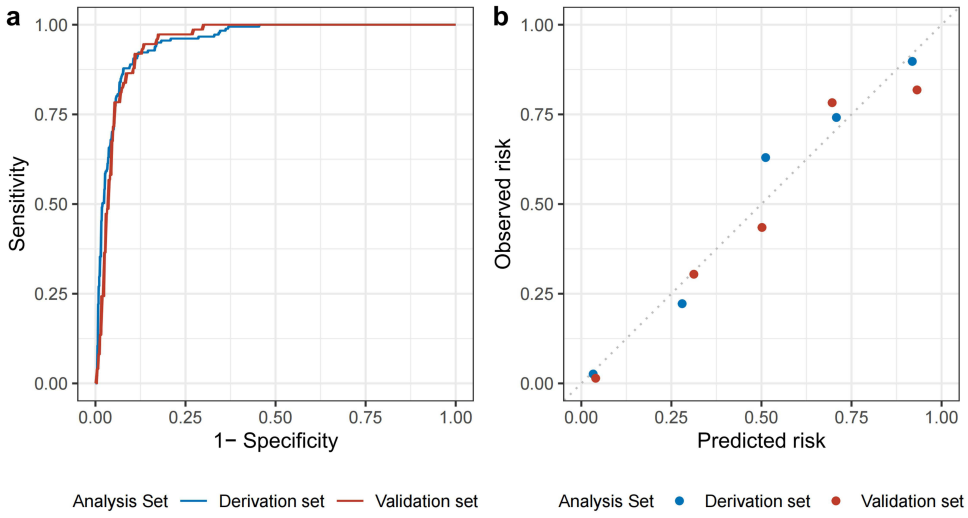

Figure 2 Prediction performance: receiver operating characteristic curve (**a**) and calibration plot (**b**).

Figure 3 A schematic representation of the clinical usefulness of the prediction model. Red circles with "G" denote GHD patients, while blue circles with "N" denote non-GHD patients. Yellow borders indicate predictions of GHD, while green borders indicate predictions of non-GHD.

A schematic representation of the clinical usefulness of the model is shown in Figure 3. Briefly, out of 100 patients with short stature, 23 patients were actually GHD, while 77 patients were actually non-GHD. Among the 23 GHD patients, 22 were predicted as GHD, while 1 was falsely predicted as non-GHD. Among the 77 non-GHD patients, 66 were predicted as non-GHD, while 11 were falsely predicted as non-GHD and had unnecessary GH stimulation tests. Under this scenario, 33 patients predicted to have GHD would undergo GH stimulation tests, which is far better than performing the tests for all 100 patients.

However, it is important to acknowledge the risk that a patient with true GHD may be overlooked (a false negative) using this screening model. Missing a diagnosis of GHD can lead to delayed or suboptimal treatment, potentially impacting final adult height, pubertal development, bone mineral density, and metabolic health. Therefore, while our model serves as an efficient primary screening tool, clinicians should maintain high clinical vigilance. This includes considering repeat assessments if growth concerns persist despite a negative screening result, and ensuring that clinical judgment always supersedes a single model prediction. Regular follow-up and re-evaluation based on evolving clinical signs and symptoms are essential to mitigate the risk of delayed diagnosis and ensure optimal patient outcomes.

Most previous studies for GHD screening were univariate, investigating the feasibility of using a specific clinical characteristic as a diagnostic marker of GHD. The diagnostic markers identified include clinical features of IGF-1, IGFBP-3, and the IGF-1 to IGFBP-3 ratio,[13–17] as well as MRI features including pituitary volumes.[19,20] However, the performance can be enhanced using multivariate approaches. Only a limited number of studies have reported prediction models for diagnosing GHD, and their performance ranges from weak to moderate. A prediction model developed in a cohort of 1496 Chinese short stature children reports poor discrimination with an AUC of 0.593.[21] This urges the need

https://doi.org/10.2147/JMDH.S534760

to develop more accurate prediction models for GHD screening. Our model addresses this need and greatly advances GHD screening accuracy. The model demonstrated consistently good discrimination, with an AUC of 0.952 (95% CI 0.937, 0.967) in the derivation set and 0.950 (95% CI 0.927, 0.973) in the validation set. Furthermore, it showed high overall prediction accuracy, with a sensitivity of 0.895 and a specificity of 0.895 in the derivation set, which were 0.946 and 0.851, respectively, in the validation set. The model also consistently demonstrated reliable calibration in both analysis sets, suggesting adequate performance for clinical use.

However, we noticed that the error profiles are not the same between the two analysis sets. While the discrimination and calibration are similar, the model is more specific in the derivation set and more sensitive in the validation set. This may be a result of the selected threshold not being perfect in the validation set, indicating that the threshold should be tuned for each specific patient population in different clinical settings.

To assess the consistency of our model's performance, we conducted stratified analyses by sex and age subgroups, as detailed in Table S2. The model demonstrated consistently strong discriminative ability across both male and female subgroups, with AUCs ranging from 0.933 to 0.974 in the validation set. Similarly, performance remained robust across the age subgroups (3–5 years and 6–12 years), with validation AUCs ranging from 0.933 to 0.975. The model maintains good performance across these key demographic strata.

Our final model includes four predictors. Younger age, delayed bone age, lower IGF-1 SDS, and a lower IGF-1/IGFBP-3 ratio were shown to be risk factors for GHD. Biomarkers of the GH-IGF-1 axis, specifically lower IGF-1 SDS and a lower IGF-1/IGFBP-3 ratio, are central to GHD diagnosis. IGF-1, the primary mediator of growth hormone action, has levels directly reflecting GH secretion, making low IGF-1 SDS indicative of GH insufficiency.[26] The IGF-1/IGFBP-3 ratio adds further insight, as IGFBP-3 is the major binding protein for IGF-1, regulating its bioavailability.[27] While IGF-1 levels can be influenced by factors like nutrition and liver function, the ratio provides a more stable measure of GH activity by accounting for variations in binding proteins.[15] In GHD, the deficiency in GH often leads to a delay in bone maturation, resulting in a bone age that is significantly younger than the child's actual age. This delay is a clinically recognized marker of disrupted growth and development, reflecting the anabolic actions of GH on bone.

These findings are largely consistent with a previous study comparing the clinical characteristics of GHD and non-GHD children, except for the age, which was similar in that study.[15] Comparisons of age between GHD and non-GHD patients conducted in different studies showed different results. In the previous model,[21] age was a risk factor for GHD. A Chinese study exploring the etiology and clinical characteristics of short stature children also revealed that the age at diagnosis of GHD patients is comparatively older than that of ISS patients.[28] Conversely, a Korean study comparing GH treatment in GHD and ISS patients revealed that GHD patients are at a younger age when starting GH treatment.[29] The results are corroborated by studies in Israel and Spain.[30,31] Variations in the study populations, enrolment process, and selection criteria may account for the contradiction.

These observed inconsistencies, alongside other critical factors, may undermine the generalizability of our prediction model. Firstly, it is important to acknowledge that the definition of GHD, specifically the peak GH threshold used in stimulation tests, is not universally standardized. A recent review highlighted the considerable uncertainty in diagnosing GHD using stimulated peak GH levels, noting that proposed diagnostic thresholds range from 5 to 10 μg/L and emphasizing that none of these cut-off values are supported by strong evidence.[32] This variability undermines the consistency of the gold standard across different studies and clinical settings. Our study adopted a cut-off value of 10 μg/L, which is consistent with clinical guidelines in China.[24] Therefore, while our model demonstrates robust performance based on the established criteria, its application in regions adhering to different GH thresholds would require careful consideration and potentially recalibration or re-evaluation to ensure optimal performance.

Secondly, variations in assay methods for IGF-1 and IGFBP-3 across laboratories can lead to different absolute values, even for the same biological sample, potentially affecting the performance of models relying on these biomarkers.[33] Thirdly, ethnicity-specific growth patterns can influence anthropometric measures and their interpretation.[34] Given these factors, external validation of our model is needed to prove its applicability and performance in diverse clinical settings and patient populations.

In our study, we excluded 80 patients with causes of short stature other than GHD, ISS, SGA, or genetic syndromes due to their low incidence rates. Specifically, this included 16 patients with sexual development disorders, 34 with pituitary conditions, 15 with thyroid disorders, and 15 with chronic systemic diseases. This exclusion was a pragmatic choice aimed at enhancing the model's discriminative power to distinguish GHD from the most common causes of short stature, aligning with its intended use as a primary screening tool. Including conditions with very low prevalence could lead to statistical instability and dilute the model's ability to accurately differentiate the most clinically relevant groups for initial screening. While this approach introduces a degree of spectrum bias, future research should explore the model's utility in broader, unselected pediatric populations or develop more comprehensive models that incorporate these less common but clinically important differential diagnoses.

While our prediction model enhances GHD screening for a defined subset of short stature patients, we acknowledge the broader and complex etiopathogenesis of short stature. Beyond primary GHD, genetic factors vary widely; some syndromes (eg, Noonan) benefit from GH therapy even without classical GHD.[35] Additionally, secondary GHD from acquired conditions, such as intracranial lesions, presents distinct diagnostic challenges.[36] Clinicians must remain vigilant for this broader spectrum of etiologies, as these complex cases may require investigations beyond our model's screening scope.

To enhance the precision and clinical applicability of our GHD prediction model, we focused on children with eutrophic short stature by excluding those with a BMI outside the 15th to 85th percentile range. This exclusion was crucial, as both underweight and overweight/obesity can significantly confound the interpretation of GH-related parameters and potentially lead to misdiagnosis.[37,38]

We acknowledge several limitations impacting the cautious interpretation of our findings. As a single-center, retrospective study, our population may lack broad representativeness, introducing selection bias from specific referral patterns and diagnostic practices. The retrospective nature also carries inherent risks of information bias (eg, misclassification due to varied data recording) and challenges in fully accounting for all potential confounding factors, as historical data collection relies on prior completeness and accuracy. Finally, the model's current reliance on a computer program for score calculation may limit immediate widespread adoption. Despite these, our model demonstrates robust performance within its derivation and validation cohorts.

Conclusion

In conclusion, we have developed an accurate prediction model for screening GHD in Chinese children with short stature. This model has the potential to serve as a non-invasive, primary screening and triage tool for GHD in children presenting with short stature in clinical settings. By identifying individuals at higher risk, our model may help optimize the diagnostic pathway by prioritizing those who may benefit most from gold-standard GH stimulation tests, thereby reducing unnecessary invasive procedures. Further prospective validation and implementation studies are warranted to confirm its utility.

Funding

This work was supported by Shaanxi Provincial Key Research and Development Project (No. 2023-YBSF-065), Xi'an Science and Technology Project (21YXYJ0013) and Xi'an Children's Hospital Research Project (No. 2022F04).

Disclosure

The authors report no conflicts of interest in this work.

References

1. Wheeler PG, Bresnahan K, Shephard BA, Lau J, Balk EM. Short stature and functional impairment. *Arch Pediatr Adolesc Med*. 158. 3:236 doi: 10.1001/archpedi.158.3.236
2. Backeljauw P, Cappa M, Kiess W, et al. Impact of short stature on quality of life: a systematic literature review. *Growth Horm IGF Res*. 2021;57–58:101392. doi:10.1016/j.ghir.2021.101392
3. Baron J, Sävendahl L, De Luca F, et al. Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol*.;11(12):735–746. doi:10.1038/nrendo.2015.165

4. Collett-Solberg PF, Jorge AAL, Boguszewski MCS, et al. Growth hormone therapy in children; research and practice – a review. *Growth Hormone IGF Res*. 2019;44:20–32. doi:10.1016/j.ghir.2018.12.004
5. Labarta JI, Ruiz JA, Molina I, De Arriba A, Mayayo E, Longás AF. Growth and growth hormone treatment in short stature children born small for gestational age. *Pediatr Endocrinol Rev*. 2009;6 Suppl 3(3):350–357.
6. Rani D, Shrestha R, Kanchan T, Krishan K. Short stature. 2023.
7. Inzaghi E, Reiter E, Cianfarani S. The challenge of defining and investigating the causes of idiopathic short stature and finding an effective therapy. *Horm Res Paediatr*. 2019;92(2):71–83. doi:10.1159/000502901
8. Wit JM, Clayton PE, Rogol AD, Savage MO, Saenger PH, Cohen P. Idiopathic short stature: definition, epidemiology, and diagnostic evaluation. *Growth Horm IGF Res*. 2008;18(2):89–110. doi:10.1016/j.ghir.2007.11.004
9. Haghshenas Z, Sotoudeh K, Karamifar H, Karamizadeh Z, Amirhakimi G. The role of insulin like growth factor (IGF)-1 and IGF-binding protein-3 in diagnosis of growth hormone deficiency in short stature children. *Indian J Pediatr*. 2009;76(7):699–703. doi:10.1007/s12098-009-0115-0
10. Grimberg A, DiVall SA, Polychronakos C, et al. Guidelines for growth hormone and insulin-like growth factor-i treatment in children and adolescents: growth hormone deficiency, idiopathic short stature, and primary insulin-like growth factor-i deficiency. *Horm Res Paediatr*. 2016;86 (6):361–397. doi:10.1159/000452150
11. Albrecht A, Penger T, Marx M, Hirsch K, Dörr HG. Short-term adverse effects of testosterone used for priming in prepubertal boys before growth hormone stimulation test. *J Pediatr Endocrinol Metab*. 31. 1:21–24 doi: 10.1515/jpem-2017-0280
12. Yau M, Rapaport R. Growth hormone stimulation testing: to test or not to test? That is one of the questions. *Front Endocrinol*. 2022;13:902364. doi:10.3389/fendo.2022.902364
13. Ranke MB. Insulin-like growth factor binding-protein-3 (IGFBP-3). *Best Pract Res Clin Endocrinol Metab*. 2015;29(5):701–711. doi:10.1016/j.beem.2015.06.003
14. Giannakopoulos A, Efthymiadou A, Chrysis D. Insulin-like growth factor ternary complex components as biomarkers for the diagnosis of short stature. *Eur J Endocrinol*. 185. 5:629–635 doi: 10.1530/EJE-21-0475
15. Haj-Ahmad LM, Mahmoud MM, Sweis NWG, et al. Serum IGF-1 to IGFBP-3 molar ratio: a promising diagnostic tool for growth hormone deficiency in children. *J Clin Endocrinol Metab*.;108(4):986–994. doi:10.1210/clinem/dgac609
16. Inoue-Lima TH, Vasques GA, Nakaguma M, et al. A bayesian approach to diagnose growth hormone deficiency in children: insulin-like growth factor type 1 is valuable for screening and IGF-binding protein type 3 for confirmation. *Horm Res Paediatr*. 2020;93(3):197–205. doi:10.1159/000509840
17. Boquete HR, Sobrado PGV, Fideleff HL, et al. Evaluation of diagnostic accuracy of insulin-like growth factor (IGF)-I and IGF-binding protein-3 in growth hormone-deficient children and adults using ROC plot analysis. *J Clin Endocrinol Metab*. 2003;88(10):4702–4708. doi:10.1210/jc.2003-030412
18. Shen Y, Zhang J, Zhao Y, Yan Y, Liu Y, Cai J. Diagnostic value of serum IGF-1 and IGFBP-3 in growth hormone deficiency: a systematic review with meta-analysis. *Eur J Pediatr*. 2015;174(4):419–427. doi:10.1007/s00431-014-2406-3
19. Kessler M, Tenner M, Frey M, Noto R. Pituitary volume in children with growth hormone deficiency, idiopathic short stature and controls. *J Pediatr Endocrinol Metab*. 29. 10. doi:10.1515/jpem-2015-0404
20. Han X, Xiu J, Huang Z, et al. Three-dimensional magnetic resonance volumetry of the pituitary gland is effective in detecting short stature in children. *Exp Ther Med*. 2014;8(2):551–556. doi:10.3892/etm.2014.1778
21. Li XJ, Luo FH, Pei Z, et al. Analyzing the etiology of 1496 children with short stature and establishing the predict model of growth hormone deficiency using IGF-1 levels. *J Clin Pediatr*. 2012;30(12):1110–1115.
22. Cong M, Qiu S, Li R, Sun H, Cong L, Hou Z. Development of a predictive model of growth hormone deficiency and idiopathic short stature in children. *Exp Ther Med*. 21. 5:494 doi: 10.3892/etm.2021.9925
23. Li H, Ji CY, Zong XN, Zhang YQ. Height and weight standardized growth charts for Chinese children and adolescents aged 0 to 18 years. *Zhonghua Er Ke Za Zhi*. 2009;47(7):487–492.
24. Subspecialty Group of Endocrinology H and MD the S of PCMA. [Chinese guidelines for the diagnosis and treatment of pediatric growth hormone deficiency]. *Zhonghua Er Ke Za Zhi*. 62. 1. 5–11. doi:10.3760/cma.j.cn112140-20230914-00183
25. Cao B, Peng Y, Song W, et al. Pediatric continuous reference intervals of serum insulin-like growth factor 1 levels in a healthy chinese children population – based on PRINCE study. *Endocr Pract*. 2022;28(7):696–702. doi:10.1016/j.eprac.2022.04.004
26. Ibba A, Corrias F, Guzzetti C, et al. IGF1 for the diagnosis of growth hormone deficiency in children and adolescents: a reappraisal. *Endocr Connect*. 2020;9(11):1095–1102. doi:10.1530/EC-20-0347
27. Yamada PM, Mehta HH, Hwang D, Roos KP, Hevener AL, Lee KW. Evidence of a role for insulin-like growth factor binding protein (IGFBP)-3 in metabolic regulation. *Endocrinology*. 151. 12:5741–5750 doi: 10.1210/en.2010-0672
28. Han X, Dong Z, Zhang W, et al. Analysis of the etiology and clinical characteristics of short stature. *J Clin Pediatr*. 2019;37(1):39–42.
29. Kim SA, Choe YR, Yang EM, Kim CJ. Comparison of growth hormone treatment in patients with idiopathic short stature and idiopathic growth hormone deficiency. *Chonnam Med J*. 2014;50(2):63. doi:10.4068/cmj.2014.50.2.63
30. Ben-Ari T, Chodick G, Shalev V, Goldstein D, Gomez R, Landau Z. Real-world treatment patterns and outcomes of growth hormone treatment among children in israel over the past decade. *Front Pediatr*. 2004–2015. 9. doi:10.3389/fped.2021.711979
31. Ariza-Jimenez AB, Leiva Gea I, Martinez-Aedo Ollero MJ, Lopez-Siguero JP. Isolated growth hormone deficiency and idiopathic short stature: comparative efficiency after growth hormone treatment up to adult height. *J Clin Med*. 10. 21:4988 doi: 10.3390/jcm10214988
32. Yuen KCJ, Johannsson G, Kky H, Miller BS, Bergada I, Rogol AD. Diagnosis and testing for growth hormone deficiency across the ages: a global view of the accuracy, caveats, and cut-offs for diagnosis. *Endocr Connect*. 12. 7. doi:10.1530/EC-22-0504
33. Mohammed-Ali Z, Delaney S, Singh R, et al. Bias in IGF-1 concentrations and interpretation across three different clinical laboratory assays. *Clin Biochem*. 2022;108:14–19. doi:10.1016/j.clinbiochem.2022.06.009
34. Sacker A, Kelly YJ. Ethnic differences in growth in early childhood: an investigation of two potential mechanisms. *Eur J Public Health*. 22. 2:197–203 doi: 10.1093/eurpub/ckr032
35. Sodero G, Cipolla C, Pane LC, et al. Efficacy and safety of growth hormone therapy in children with Noonan syndrome. *Growth Hormone IGF Res*. 2023;69–70:101532. doi:10.1016/j.ghir.2023.101532

36. Sodero G, Cipolla C, Camporesi A, et al. Endocrinologic dysfunctions and neuropsychiatric sequelae in pediatric patients with a history of central nervous system infection (ENDLESS): a prospective monocentric study. *Pediatr Infect Dis J*.;44(4):310–317. doi:10.1097/INF.0000000000004645
37. Stanley TL, Levitsky LL, Grinspoon SK, Misra M. Effect of body mass index on peak growth hormone response to provocative testing in children with short stature. *J Clin Endocrinol Metab*. 94. 12:4875–4881 doi: 10.1210/jc.2009-1369
38. Fazeli PK, Klibanski A. Determinants of GH resistance in malnutrition. *J Endocrinol*. 2014;220(3):R57–65. doi:10.1530/JOE-13-0477